

# РАЗРАБОТКА БИБЛИОТЕКИ НЕЙРОСЕТЕВОГО РАСПОЗНАВАНИЯ РУКОПИСНЫХ СИМВОЛОВ НА МАШИНОПЕЧАТАЕМЫХ БЛАНКАХ

*Н. Г. Авдеев, аспирант гр. А0-39  
Томский политехнический университет  
E-mail: ndb1@tpu.ru*

## Введение

Существующий формат проведения ЕГЭ и ОГЭ ставит перед организаторами задачу анализа информации, представленной в виде изображений – произвести так называемое off-line распознавание уже написанного на бумаге текста. Задача обработки и распознавания изображений относится к разряду трудно формализуемых задач и является одной из наиболее важных на сегодняшний день.

## Выбор технологии распознавания

Для поставленной задачи были выбраны сверточные нейронные сети. Выбор сверточных нейронных сетей обусловлен следующими факторами:

- Начиная с 2012 года сверточные сети занимают первые места на конкурсе ImageNet [1];
- Сверточные нейронные сети устойчивы к аффинным преобразованиям;
- Сверточные нейронные сети учитывают топологию изображений [2].

## Подготовка обучающей и валидационной выборок

Обучающая выборка была составлена из бланков репетиционного ЕГЭ. Из бланков были вырезаны изображения символов размером 64x64. В ходе подготовки обучающей выборки было выявлено, что некоторые символы в ответах участников встречаются достаточно редко (буквы «Ё», «Ф», «Ц», «Щ», «Ъ», «Э» и «Ю»). Изображения различных классов должны быть представлены в обучающей выборке примерно в одинаковых пропорциях. Если один из классов будет преобладать, то это может привести к «перекосу» в процессе обучения модели [3]. Было решено восполнить недостаток в данных классах с помощью метода эластичных искажений [4]. В результате в каждом классе подготовленной выборки количество символов стало равным примерно 6000.

## Реализация метода распознавания на основе свёрточной нейронной сети

Для программы репетиционного экзамена было решено использовать две нейронные сети для распознавания следующих множеств символов:

- цифры, минус, запятая;
- символы русского алфавита.

Использование 2-х нейронных сетей необходимо для того, чтобы избежать неоднозначности в распознавании символов (цифра «0» и буква «О», цифра «3» и буква «з» и т. д.).

Были проведены эксперименты по подбору архитектуры нейронной сети. Количество слоев свертки варьировалось в пределах от 2 до 4, размер ядра свертки от 3x3 до 9x9, количество карт признаков на 1 слое – от 5 до 40. Количество карт признаков на каждом последующем слое удваивается. В подвыборочном слое использовалась операция MaxPooling с размером окна 2 и шагом 2. Для ускорения процесса обучение производилось на выборке, полученной следующим образом: из исходной выборки для каждого класса выбиралось 100 случайных букв для обучения и 10 случайных букв для валидации результата. Было проведено 100 итераций обучения. Результирующее значение точности распознавания было получено путем вычисления среднего арифметического из значений точности распознавания, полученных на каждой итерации.

На первом этапе эксперимента было протестировано 30 архитектур нейронных сетей. Количество слоев свертки варьировалось в пределах от 2 до 4, размер ядра свертки от 3x3 до 9x9, количество карт признаков на 1 слое оставалось равным 20. Архитектура выбиралась таким образом, чтобы не происходила потеря информации с изображения: если на одном из выходов сверточного слоя ширина или высота выходного изображения получалась нечетной, то при применении операции MaxPooling на подвыборочном слое крайние пиксели изображения не учитывались.

Исходя из результатов эксперимента, можно сделать вывод, что наибольшей точностью обладает нейронная сеть, имеющая архитектуру, описанную в таблице 1. Точность распознавания данной нейронной сети составляет 90.75%.

Таблица 1. Архитектура выбранной сети для распознавания букв

Номер сверточного слоя	Размер ядра свертки	Количество карт признаков
1	5x5	20
2	3x3	40
3	7x7	80
4	3x3	160

Далее для данной архитектуры экспериментально было подобрано количество карт признаков. Были проведены эксперименты по нахождению зависимости количества карт признаков от точности распознавания. В ходе эксперимента варьировалось количество карт признаков на первом слое (количество карт признаков на каждом последующем слое удваивается). Исходя из полученных результатов, было решено взять количество карт признаков на первом слое равным 25, так как дальнейшее увеличение значения количества карт признаков не ведет к большому приросту точности, а ведет лишь к увеличению времени распознавания.

Конечная архитектура продемонстрирована в таблице 2.

Таблица 2. Архитектура выбранной сети для распознавания букв

Номер сверточного слоя	Размер ядра свертки	Количество карт признаков
1	5x5	25
2	3x3	50
3	7x7	100
4	3x3	200

Аналогичный вышеописанному эксперимент был проведен для подбора архитектуры нейронной сети для распознавания цифр, минусов и запятых. Данная архитектура имеет структуру, описанную в таблице 3.

Таблица 3. Архитектура выбранной сети для распознавания цифр, минусов и запятых

Номер сверточного слоя	Размер ядра свертки	Количество карт признаков
1	5x5	25
2	3x3	50
3	3x3	100
4	5x5	200

### Разработка библиотеки распознавания

Программная часть реализована в виде dll-библиотеки, написанной на языке программирования C#. Нейронная сеть сохранена в формате .onnx. Для запуска нейронной сети используется библиотека Microsoft ONNX Runtime.

Изображение поступает на вход в виде двумерного массива float. Значению 0 в данном массиве соответствует черный цвет на изображении, а значению 1 – белый. Результатом распознавания является строка, состоящая из одного элемента, значение которого соответствует распознанному символу.

### Тестирование разработанной библиотеки

Было проведено 100 обучений нейронной сети для распознавания рукописных символов. Для каждого обучения выбиралось лучшее значение точности распознавания валидационной выборки. Результирующее значение точности распознавания было получено путем усреднения полученных значений и составило 98.4 %.

Значение точности распознавания нейронной сети для распознавания цифр, минусов и запятых было получено аналогично вышеописанному способу и составило **99.7 %**.

В силу того, что скорость работы верификатора зависит от процента верно распознанных ответов участника на бланке, имеет смысл вычислить данный показатель.

Интерфейс программы распознавания, которая использует данную библиотеку, продемонстрирован на рисунке 1.



Рис. 1. Интерфейс программы для распознавания ответов.

Для сравнения результатов распознавания полученного программного продукта с программой Abbyy FormReader, которая используется в ЦОКО Томской области на данный момент, было распознано 300 бланков русского языка и 200 бланков базовой математики. Из бланков ответов русского языка были взяты только ответы, содержащие буквы. Результаты распознавания приведены в таблице 4.

Таблица 4. Результаты распознавания

Распознаваемый класс	Количество распознаваемых ответов	Количество верно распознанных ответов программой Abbyy FormReader	Количество верно распознанных ответов разработанной программой	Точность распознавания программы Abbyy FormReader	Точность распознавания разработанной программы
Цифры, минус, запятая	3722	2656	3178	71,4%	85,4%
Русские буквы	3534	2809	2244	79,5%	63,5%

Разработанная библиотека превзошла программу Abbyy FormReader по распознаванию ответов, содержащих цифры, минусы и запятые. Полученный результат обуславливается тем, что Abbyy FormReader часто ошибочно распознает запятые как символ «^».

Разработанная библиотека показала худший результат по распознаванию ответов, состоящих из русских букв. Это обуславливается тем, что программный продукт Abbyy FormReader лучше распознает схожие по написанию буквы, а также умеет распознавать строчные буквы, которые написаны с несоблюдением образца.

#### Список использованных источников

1. ImageNet Winning CNN Architectures (ILSVRC) [Электронный ресурс] – 2020. Режим доступа: <https://www.kaggle.com/getting-started/149448> (дата обращения 05.02.2021).
2. Сверточная нейронная сет, часть 1: структура, топология, функции активации и обучающее множество [Электронный ресурс] – 2018. Режим доступа: <https://habr.com/ru/post/348000/> (дата обращения 05.02.2021).
3. Ian H. Witten, Eibe Frank and Mark A. Hall Data Mining: Practical Machine Learning Tools and Techniques. — 3rd Edition. — Morgan Kaufmann, 2011. — 664 с.
4. Patrice Y. Simard Platt Best Practices for Convolutional Neural Networks Applied to Visual Document Analysis / Patrice Y. Simard, Dave Steinkraus, John C. Platt – Proceedings of the Seventh International Conference on Document Analysis and Recognition (ICDAR 2003), 2003.